

Influence-shaping Strategy of Political Key Opinion Leaders on Social Media: An Analysis Based on Topic Modeling and Network Analysis

Jariel Yang

Introduction

In China, the influences of social media have widely permeated through people's lives and the emergence of key opinion leaders (KOLs) in cyberspace greatly strengthens the information-spreading effects via social media. This article discerns the strategies taken by political key opinion leaders (KOLs) to shape their community influences from the politics-related forum section on Weibo, one of the biggest social media in China. Based on the latent Dirichlet allocation (LDA) modeling, relevant tweets under this section were integrated to construct a tweet corpus, from which 9 topics were derived. Meanwhile, community detection was further performed to find out the top 10 most influential KOLs under the section. Combining topic modeling and community detection, the strategy that was taken by the KOLs to shape their influences was fathomed based on qualitative analyses: On one hand, KOLs particularly concentrated on ongoing political affairs related to bilateral politics with clear biases to express their opinions. On the other hand, topics that were chosen by these key opinion leaders are close to the daily lives of normal people, ensuring strong realism and understandability to immerse target users in the discussion.

Data Selection and Methods

41,848 tweets between 2011-2016 were selected for the purpose to establish a corpus for tweet topic detection, which were drawn under the politics-related forum section from Sina Weibo, one of the biggest Chinese social media. The primary descriptive chronological analysis (Figure 1) of the cleaned 11,558 tweet data presents that the number of tweets posted in 2011 ranked the lowest. In 2012, increasing texts were posted on this social media and fluctuated a bit from 2013 to 2014. There appeared a surge in the number of tweets posted in 2015 and 2016. After conducting the statistical description, I performed natural language processing depending on the Python jiaba package, in which each word or term that primarily comprised the tweet was respectively extracted out of the original text. Meanwhile, a filtering principle was set to eliminate the words that were meaningless for the analyses, such as numbers, punctuations, particles, pronouns, quantifiers, etc. On top of this, several *stopwords* were specially chosen then added to the filtering principle according to the tweet context, such as “retweet”, “link”, etc. Those *stopwords* were of no use for the qualitative analyses because they didn’t carry concrete meanings that could be interpreted depending on the tweet context.

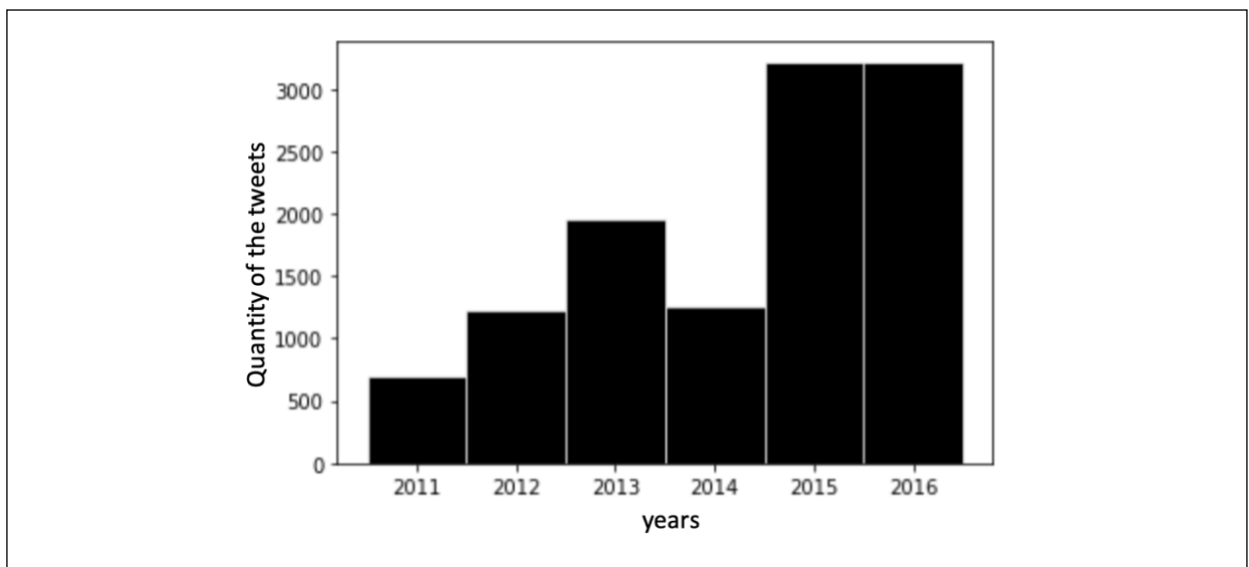


Figure 1. Yearly Distribution of Weibo Posts

After establishing the tweet corpora, the latent Dirichlet allocation (LDA) model was utilized for machine learning processing, which could assist to detect underlying topics that were discussed in the tweets (Huang 2019). According to Blei et al. (2003), the latent Dirichlet allocation (LDA) model is a generative probabilistic model for collections of discrete data such as text corpora and it is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics and each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities. With the LDA model applied to the tweet corpus, different topics can be concluded according to the probability distribution of the words that were extracted from the original tweets. Although the LDA model was primarily tapped in English, it has been proven to be effective to deal with Chinese by past practices (Lin et al., 2013). Therefore, based on the LDA model, I established several trial models with the number of topics ranging from 1 to 45 via the Python genism package, to decide which number fits the model best. In this session, perplexity and coherence were used as indicators to decide the most fitting number of topics (Huang 2019). Results (Figure 2) showed that when the number of topics was less than 15, the coherence value remained stable and the perplexity value continued to decrease. Given that setting a high number of topics might bring confusion in the interpretation process, I hereby chose 15 as the upper limit for the selection of the number of topics. Meanwhile, I chose 10 to be the lower limit of the number of topics because according to the trials, if the number of topics chosen was below 10, the final results of the topic selection would turn out to be ambiguous with different subtopics messed altogether, being hard to find out a general direction to focus on. Moreover, the coherence parameter indicated that the overall coherence of the topics was relatively high when their number was kept between 10 and 15, which further confirmed the validity of the selected range. Ultimately, I tested the topic model with the number of topics ranging from 10 to 15

respectively depending on perplexity, coherence, and the interpretability of the resultant topics, and the 13-topic model was chosen as the final model.

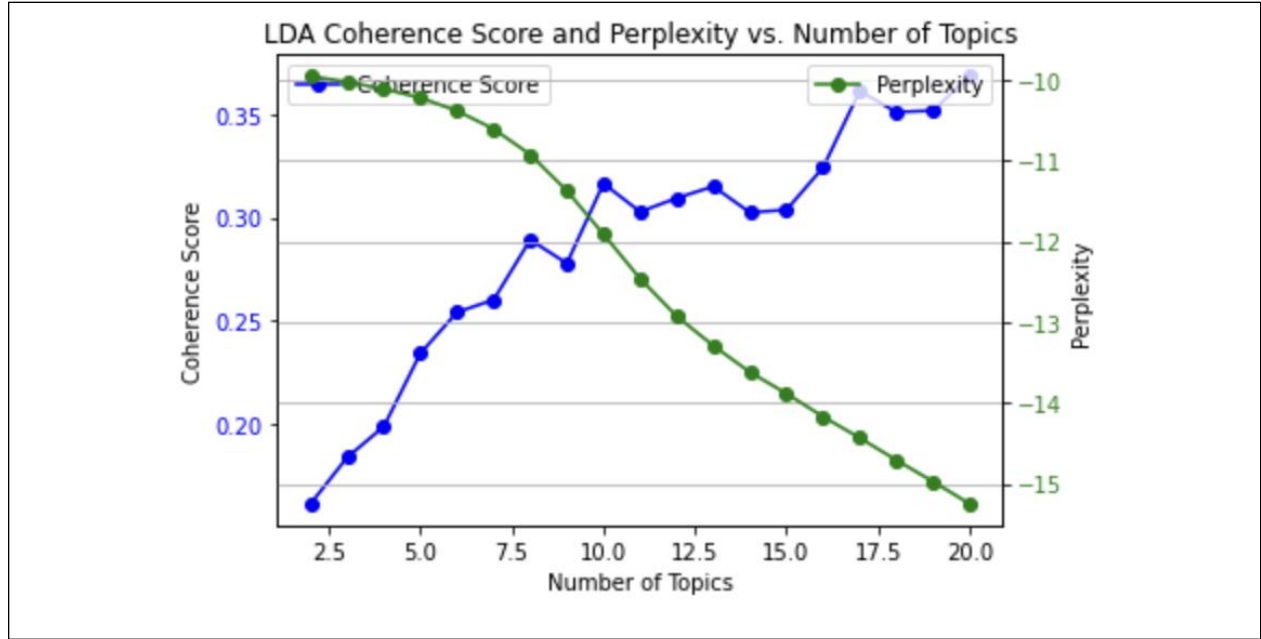


Figure 2. Model Comparisons and Selections

In addition to the topic modeling, a social network based on mutual interaction among the tweet posters was further constructed by virtue of community analysis using the Python-igraph package. Here the logic of the interactive network modeling was embodied by the “tweeting and re-tweeting” behaviors of the Weibo users: If user A reposts a tweet from user B, there will form a sort of interactive relationship between user A and user B. To simplify the social network analyses, I assumed that it’s because of the shared interests in certain topics instead of personal acquaintances that tempted user A or user B to retweet each other, thus the interactive relationship constructed between A and B was considered as a relationship without edges, i.e., there was no specific direction pointing from A to B in their relationship, or vice versa. In this sense, user A and user B are merely regarded as parts of the “audience” who are interested in the topic, and the formation of their interactive relationship is entirely driven by the topic embedded in the

discussions. Completing to establish a social network according to principles mentioned above, the component analyses concerning the social network were further conducted to find out the most representative and significant components whereas those comparatively solitary components were eliminated from the network.

Findings

Topics derived from the Weibo tweets

Based on the topic model constructed above, under each topic presented 10 keywords that were frequently discussed in the corresponding context to give a bird's eye view of the general picture of the topic. After deleting uninterpretable topics that couldn't be generalized through induction of their keywords, 9 topics were derived from the analyses, which encompass (Table 1): *U.S. Politics, Cross-Strait Relations, Chinese Military Development, China-Korea Relations, Hong Kong Problem, South China Sea Dispute, Chinese Economic Development, International Competition, and Local Society.*

Table 1. Results of topic modeling (10 most important keywords)

1.the U.S. Politics	2.Cross-Strait Relations	3.Chinese Military Development	4.China-Korea Relations	5.Hong Kong Problems
America Trump President Hillary	Taiwan Mainland Tsai Ing-wen Chinese Nationalist party	Aircraft carrier Emission Liaoning Ship First-ever	Korea Tourists Navy Samsung	Hong Kong Spokesman Independence camp Ministry of Foreign Affairs
Obama State International World Election Support	Democratic Progressive Party Media Public Straits Information Ma Ying-jeou	Fighter Plane Task Flight Air force Sky palace Fight	Park Geun Hye attack THAAD Quantity Bureau Media	Public Liberation Britain illegal Mainland Xinhua News Agency
6.South China Sea Issue	7.Chinese Economic Development	8.International Competition	9.Local Society	
South China Sea Philippine Duterte Government Police Navy U.S. China Singapore Thailand	Corporate Economy China Market Corporation Investment Million Increase Shenzhen GDP	Olympics Britain games Candidate Organization Hold Socialize Female Finals Group	Internet Suspect Elder Exposure Deny Students Families Chengdu Teachers Residents	

1. *U.S. Politics*: Topic 1 is composed of keywords such as “America”, “Trump”, “President”, “Hillary”, and “Obama”, etc. Discussions surrounding this topic mostly dwell on news regarding American politicians and the 2016 U.S. Election, which attracted attention worldwide in 2016. For instance, a tweet posted on October 13th, 2016 put: *The U.S. Election is indeed full of dramas [...] a black man who claimed to be the illegitimate child of Bill Clinton begged Hillary to accept him, saying all the people from Arkansas know this fact and his ears, mouth, jaw, teeth, eyes, and nose are pretty much the same as Clinton’s.* Besides, some tweets concentrate on the scandals of the presidential candidates such as the sexual harassment scandal of Donald Trump. Notwithstanding that some tweets under this topic were discussing American politics in a serious manner, most of them were apt to focus on scandals and gossips of the politicians.

2. *Cross-Strait Relations and Hong Kong Problem*: Topic 2 Cross-Strait Relations and topic 5 Hong Kong Problem are largely related to Chinese internal political affairs. Topic 2 refers to the relations between Mainland China and Taiwan, which is induced by keywords including “Taiwan”, “Mainland”, “Tsai Ing-wen”, “Chinese Nationalist Party”, “Democratic Progressive Party”, etc. On February 15th, 2015, Tsai Ing-wen officially declared to participate in the Taiwan Election and successfully got inaugurated on May 20th, 2016. In this period, political relations between Mainland China and Taiwan were complicated, in which the internal political relations among different parties in Taiwan were intensified while Taiwan’s external diplomatic relations with the PRC government were stagnant. The hitting public attention towards the cross-strait relations influenced lots of Weibo users to focus on this topic. For example, a tweet discussed the proposal raised by Tsai Ing-wen which was concerned with the vacation deduction of Taiwan citizens and the tweet specifically accentuated the “resentment and dissatisfaction” expressed by the public in

Taiwan. Beyond that, some tweets were exalting the preferential policies enacted by the PRC government that claimed to cancel the “boundary” between Taiwan and Mainland China by granting Taiwanese who live in Mainland China the citizen-level treatment. Topic 5 is induced by keywords such as “Hong Kong”, “Independence camp”, “Public”, “Mainland”, etc. Those keywords are highly related to the political affairs and controversy between Hong Kong S.A.R and the PRC central government, among which the conflicts between the Hong Kong independence camp and patriotists who uphold the one-China principle are clearly indicated from the tweets. For example, a tweet depicts disputes between both stances taking place on the Legislative Council of Hong Kong by means of showcasing insulting words said by the independence camp targeting the PRC government and it can be told from this tweet to have implied a strong personal bias towards the patriotist camp.

3. *China-Korea Relations and South China Sea Dispute*: Both topic 4 and topic 6 correspond to the diplomatic disputes between China and its neighboring countries. Topic 4 consists of keywords encompassing “Korea”, “Park Geun-hye”, “THAAD”, “Korean Media”, “Navy”, etc. Relevant discussions were concerned with the decision made by the Korean government to introduce the Terminal High Altitude Area Defense (THAAD) in Korea, which was perceived as an action offending the national security of China by the tweet posters. Topic 6 consists of keywords encompassing “the South China Sea”, “Philippine”, “Duterte”, etc., which are mainly discussions regarding the debates on whether the South China sea is entirely under the sovereignty of the Chinese government or not.

4. *Chinese Military and Economic Development, International Competition, and Local Society*: Topic 3 and Topic 7 respectively refer to achievements China obtained through its military construction and economic policies. The former emphasizes the Chinese army, national security,

and aviation achievements; the latter pays attention to the input and output of Chinese economic construction, which involves discussions about the financial market, local corporates, and the GDP. Topic 8 is concerned with international sporting competitions, especially the Olympic games. Keywords under topic 8 include “Olympics”, “Candidate”, “Organize”, “finals”, “Britain”, etc. Topic 9 focuses on news from Chinese local society, most of which are related to personal life events of the local Chinese, with the subtopics ranging from crimes to education.

As the analyses present, topics of the selected Weibo tweets are mainly in relation to some “hitting” international political affairs, which include public diplomacy, cross-border conflicts, etc. Meanwhile, there also involve topics regarding public discussions about the development of China itself and relevant local social events. Based on the topics concluded, to know about the chronological transforming trend of the discussion frequency of each topic, I performed statistical analysis to further delineate the changes of the topics from 2011 to 2016 (Figure 3). The monthly average discussion frequency of each topic was calculated, and results show that the transforming trend of each topic remains stable in general. Topic 6 (South China Sea Issue) shares the highest discussion frequency while that of topic 3 (Chinese Military Development) is the lowest. The discussion frequency of topic 2 (Cross-strait Relations), topic 4 (China-Korea Relations), and topic 5 (Hong Kong Problems) are approximate to one another; this situation can also be seen from the cases of topic 9 (Local Society), topic 8 (International Competition), topic 7 (Chinese Economic Development), and topic 1 (U.S. Politics), where all of them share relatively similar discussion frequencies.

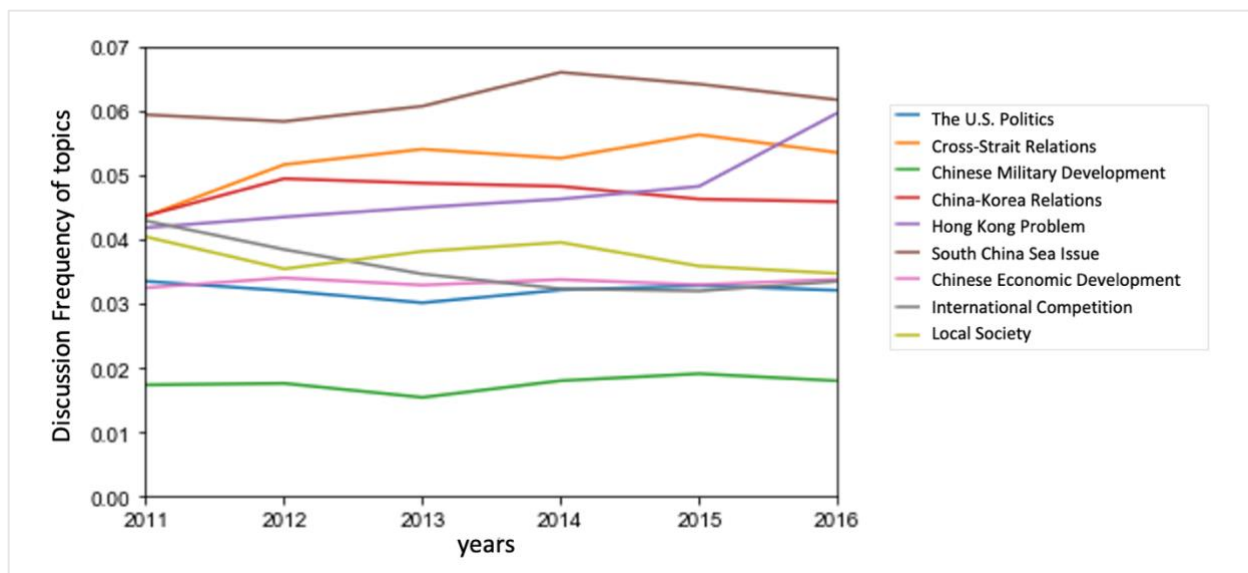


Figure 3. Transforming Trends of the Topics

Community Network Led by the Key Opinion Leaders (KOLs)

The network modeling ultimately presents an edge-free social network with 2910 users (nodes) and 6406 pairs of relationships inside, of which the overall density is 0.0015. Based on the model, community detection was performed to find out the significant components comprising this network. Depending on the component analysis method via the Python-igraph package, the biggest component that contains 1965 elements was selected from the network. After analyzing it, 25 communities were finally drawn. Those communities were visualized via Ghephi (Figure 4).

Based on the node degree values from the largest to the smallest, 10 users (nodes) were picked out to represent 10 key opinion leaders in their corresponding communities (Table 2). Top 10 key opinion leaders in the network are¹: @realGuancha, @ Tonight News Paper, @April, @Xiaodong Wang, @Fatian Wu, @Zizheng Dian, @FlyAgainstLight, @Yang Liu, @KnightontheBike, @Xiaoxin Cai. @realGuancha, for example, owns the highest degree value (1020), and its

¹ Aliases of the listed key opinion leaders were originally displayed in Chinese as: @观察者网、@今晚报、@王小明、@吴法天、@点子正、@-逆光飞行-、@刘仰、@单车上的骑士、@蔡小心

network closeness value reaches 0.534, which is the biggest among the 10 KOLs. Those values indicate that the average shortest distance (ASD) between @realGuancha and other users is the smallest, and its relationships with other users are the closest. On the contrary, the degree value (67) of @Xiaoxin Cai is the smallest among all, which implies that @Xiaoxin Cai is relatively solitary in the network and the density of its community is small compared with the other 9 KOLs. This situation is understandable: the previous topic model reveals that discussions of the Weibo tweets are mainly concerned with political affairs, indicating users in this Weibo section are mostly interested in topics related to politics. Plus, @realGuancha is one of the most renowned leading media in China that covers a wide range of topics regarding political and social events, therefore the social network on which it centers will undoubtedly be the largest. Apart from @realGuancha, other KOLs also used to post tweets related to political affairs.

To sum up, the interactions under the forum section between KOLs and other users are mostly concerned with politics-related topics such as political aptitudes, diplomatic strategy, military action, etc., in which the KOLs typically play a role in transmitting information and shaping their personal influences by imbuing opinions or values with a strong bias to their followers.

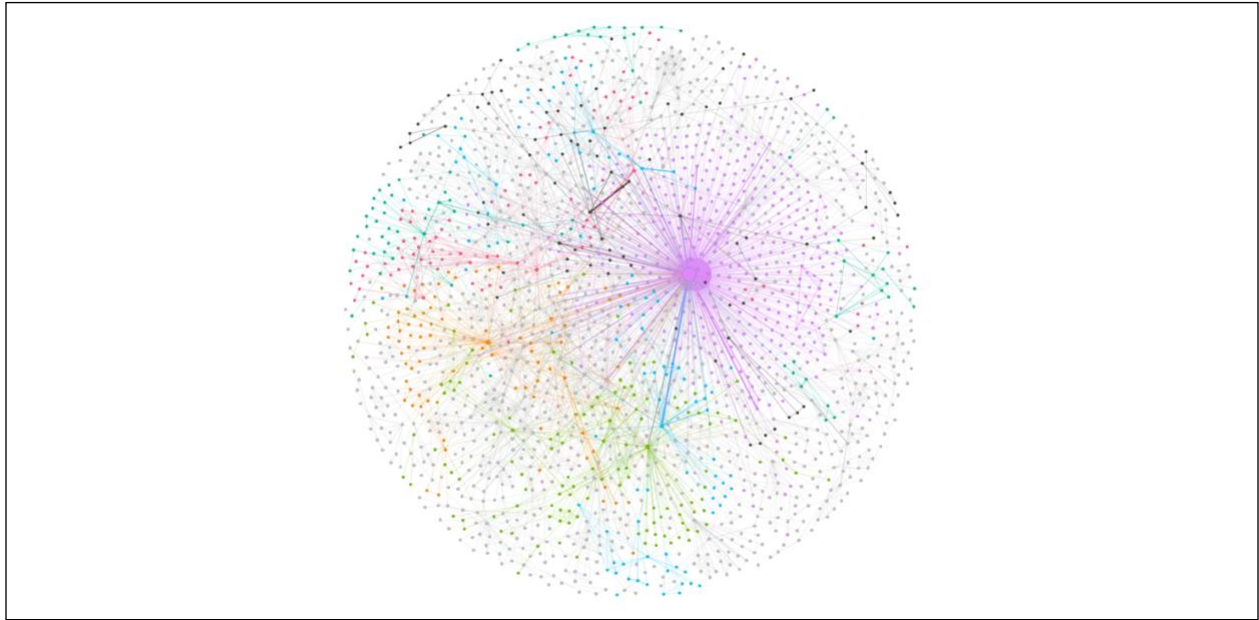


Figure 4. Community Detection of the Social Network

Table 2. Top 10 Key Opinion Leaders Selection

Ranking	Alias	Degree	Network closeness
1	@realGuancha	1020	0.534
2	@ Tonight News Paper	111	0.381
3	@April	105	0.384
4	@Xiaodong Wang	103	0.388
5	@Fatian Wu	91	0.382
6	@Zizheng Dian	85	0.380
7	@FlyAgainstLight	81	0.376
8	@Yang Liu	76	0.380
9	@KnightontheBike	71	0.375
10	@Xiaoxin Cai	67	0.376

Community Influencer Shaping Strategy: Controversy and Realisticity

Social media influencers typically seek to establish a fixed pattern that's matched to their personal figures to activate discussions around themselves. In most cases, the fixed pattern refers

to a common topic with shared interests and attention in the community (Zhang et al. 2016). In this regard, to fathom the strategy taken by the KOLs in terms of shaping their influences, I investigated topics that were frequently discussed in each community. Based on the topic model, I respectively calculated the average discussion frequency of the 9 topics in each of the 10 communities selected. The results (Figure 6) show that there is no apparent difference in the topics that were discussed in each community and the discussion frequency distribution of the 9 topics in each community is basically in line with that of the total communities. By the average discussion frequency from the largest to the smallest, the topics can be ranked as (1) *South China Sea Issue*, (2) *Cross-strait Relations*, (3) *Hong Kong Problems*, (4) *China-Korea Relations*, (5) *Local Society*, (6) *International Competition*, (7) *Chinese Economic Development*, (8) *American Politics*, and (9) *Chinese Military Development*. On top of this, I further categorized those topics into 3 dimensions: *Bilateral Politics*, which includes (1)-(4); *Society and People*, which includes (5) and (6); *Unilateral Politics*, which includes (7)-(9).

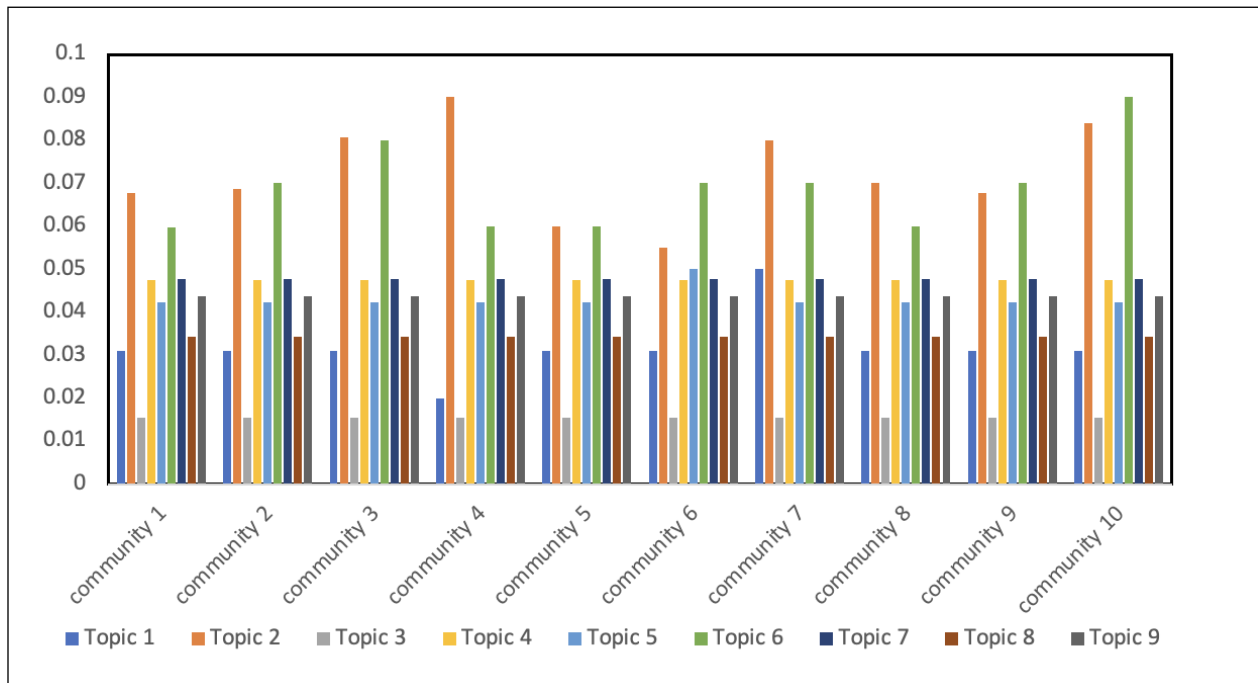


Figure 6. Distribution of Discussion Frequency in Each Community

Depending on the qualitative analyses through the combination of the topics, community KOLs, and their original tweet content, it can be found out that political key opinion leaders on Weibo prefer to rely on topics with high controversy and strong realisticity to shape their influences, in which their biases and political aptitudes should be clearly presented.

Firstly, topics related to “Bilateral Politics” share the highest discussion frequency among the 10 communities. This is because topics in this category are closely aligned with the ongoing political affairs, in which each of them clearly presents two or more antithetical stances – South China Sea Issue, for example, is mainly concerned with the dispute between China and its neighboring countries, where different sides were holding conflicting viewpoints on the sovereignty of South China Sea. In order to make good use of the “event controversy”, political key opinion leaders tend to construct their virtual figures with a strong (or even extreme) bias and a clear aptitude. This can be glimpsed from an interaction between @Xiaodong Wang and other users:

Bravo! //@Xiaodong Wang: My attitude is super clear and that it isn't matched to either side mentioned by you doesn't mean I don't have a clear stance. I will always stand for China but as for you, I just don't want to cater to either of the sides. I think you guys are pieces of shit! Come and bite me, can't you? //@DDZGG: Since we are talking about politics, our attitude needs to be clear. It's incorrect to switch back and forth between the left-wing and the right-wing. Also, it's impossible to cater to both sides, which will only lead to a bad ending. Xiaodong Wang is an example of this, it's said by Wiki that he has met with the U.S. officials for 4 times secretly and he was called “contact” by them. Besides, Xiaodong Wang once revealed that Chinese Army was going to take military action in Xinjiang, in which he said that the source of the information came from his contact in the army. Wang said he also supported democracy, human rights, and reform,

however, his words that were in support of democracy and criticized Zedong Mao were deleted in his documentary.

The interactive tweet here involves topics related to political relations between China and the U.S., where the conversation between @Xiaodong Wang and other users carries acute conflicts. A user repatriated @Xiaodong Wang because he thought @Xiaodong Wang had a bias leaning toward the U.S. rather than China. Instead of taking a mild strategy, @Xiaodong Wang took an extreme approach in response to the critiques, with some cursive words used and an anxious mood expressed in the tweet. However, the extreme response from @Xiaodong Wang didn't infuriate his followers. Rather, most followers were for @Xiaodong Wang, and this response was even complimented with words like "*Bravo!*" by some users. This conversation clearly presents the strategy taken by the KOLs to shape their influences: The acuter, the better.

In addition to basing the discussion topics on some controversial affairs, the ideal topics should also address "realisticity" by closely relating themselves to the daily lives and events of normal people thus enabling general users to engage in the discussion. This is the reason why the topics under "Society and Life" obtained high discussion frequency. Take an interactive tweet from @realGuancha and its followers as an example:

@LYWX: Though it's inappropriate for the public to force rich people to donate money to charity, rich people in China still need to be more benevolent and responsible. Because they are not living in a virtual world! //@realGuancha: People earned their money themselves, and rich people are not obligated to donate money reluctantly. Donating is an individual behavior based on consent rather than a kind of public obligation. Therefore, people shouldn't force anyone who seems to be rich to donate money to charity.

This tweet mainly discussed whether people with a higher socioeconomic status should take more responsibility for the public good. This is a realistic topic because the background behind the discussion is based on an explosion accident taking place in Tianjin China, which caused great losses to the city. After the accident, some people flooded into Weibo to force Jack Ma, one of the richest men in China, to donate money to the reconstruction project in Tianjin. This event had appealed wide attention from the public by virtue of its close relation to their daily lives and an underlying social puzzle that everyone could comment on.

As for the final category “Unilateral Politics”, it serves more as the comparison to the prior two categories as it’s less controversial and is far away from the daily lives of normal people. The topics under “Unilateral Politics” tend to involve information of a single country regarding its politics, economy, and military construction, which are things that have already reached a consensus on. In the meantime, fully understanding those topics requires individuals to master some academic knowledge. Therefore, topics under “Unilateral Politics” generally gain less public attention on Weibo, becoming topics that are least chosen by the political KOLs to shape their influences.

To summarize, qualitative analyses based on the topic model and community detection point out that political key opinion leaders tend to depend on topics with high controversy and strong realism to shape their influences in the communities. On one hand, they particularly concentrated on ongoing political affairs related to bilateral politics with clear biases to express their opinions. On the other hand, topics that were chosen by these key opinion leaders are typically close to the daily lives of normal people, ensuring strong realism and understandability to immerse target users in the discussion.

References:

- Du, Y., et al. (2021). "Detection of key figures in social networks by combining harmonic modularity with community structure-regulated network embedding." *Information Sciences* 570: 722-743.
- Yong Gui, R. H., Yi Ding (2020). "Three faces of the online leftists: An exploratory study based on case observations and big-data analysis." *Chinese Journal of Sociology* 6(1): 67-101.
- Li, C., et al. (2019). "Opinion community detection and opinion leader detection based on text information and network topology in cloud environment." *Information Sciences* 504: 61-83.
- Huang, R. (2019). "Network fields, cultural identities and labor rights communities: Big data analytics with topic model and community detection." *Chinese Journal of Sociology* 5(1): 3-28.
- Zhian Zhang, Y. H. (2016). "Individual Emotive, Social Emotion, and Collective Will: A Study on Irrationality and Its Factors." *Shanghai Journalism Review* 11: 16-22.
- Ronggui Huang, Y. G., Xiaoyi Sun (2015). "Inter-organizational network structure and formation mechanisms in Weibo space: A study of environmental NGOs." *Chinese Journal of Sociology* 1(2): 254-278.
- Lin Ping, H. W. (2013). "Event Topic Evolution of Network Public Opinions: An Analysis Based on LDA Model." *JOURNAL OF INTELLIGENCE* 32(12): 26-30.
- Jordan, D. M. B. A. Y. N. M. I. (2003). "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993-1022.